AWRA 2010 SUMMER SPECIALTY CONFERENCE
San Juan, Puerto Rico
August 30– September 1, 2010

RAINFALL ESTIMATION OVER PUERTO RICO USING A CLASSIFICATION SYSTEM AND ANN

Melvin J. Cardona-Soto, Nazario D. Ramirez-Beltran, and Eric W. Harmsen*

ABSTRACT: Precise remote detection and estimation of rainfall has become critical for protecting human lives and infrastructure. Researchers have developed diverse algorithms for deriving rainfall rates from instruments on geostationary satellite platforms such as the Geostationary Operational Environmental Satellites (GOES) due to its relatively high spatial and temporal resolution and uniform spatial coverage.

Validations of the operational NOAA/NESDIS Hydro-Estimator (HE) algorithm conducted over Puerto Rico (PR) at a satellite pixel and island-wide scale showed that the algorithm has a low probability of detection. The poor performance of the HE over PR may in part be due to the fact that the algorithm was designed to operate over the continental United States and conditions over PR are considerably different. In order to achieve greater accuracy of detection and estimation over PR, a new rainfall algorithm is under development. The algorithm utilizes data from multiple bands of GOES-12 to extract diverse features from clouds (e.g., Brightness Temperature, Visual Reflectance, and Albedo). These features are utilized to perform a supervised classification of the image pixels into 4 previously defined classes. The characterized classes will only provide rainfall detection information. After the classification is completed, two artificial neural networks will be utilized to find a feature-rain rate relationship for each class. Preliminary results in terms of rainfall detection show that the algorithm's classification system has great potential for outperforming the HE over PR.

KEY TERMS: rainfall estimation algorithm; cloud classification system; Neural Networks; Puerto Rico; Hydro-Estimator; remote sensing

INTRODUCTION

Correct rainfall detection and estimation are of interest and crucial for various metrological and hydrological applications (Vila and Velasco, 2002). Having correct information in a timely manner, may help prevent catastrophic events that may harm infrastructure and human lives. Over areas where rain gauges are scarce, remotely-derived quantitative precipitation estimates (QPE) are extremely important (Vila et al., 2003).

Throughout the years researchers have developed diverse algorithms to detect and derive rainfall from remotely sensed imagery over continental areas. Some of these algorithms are: (1) GOES Precipitation Index (GPI) (Arkin and Meinsner, 1987), (2) Auto Estimator (AE) (Vicente et al., 1998), (3) Precipitation Estimation from Remotely Sensed Information using an Artificial Neural (PERSIANN) (Sorooshian et al., 2000), (4) GOES Multi-Spectral Rainfall Algorithm (GMSRA) (Ba and Gruber, 2001), (5) Hydro Estimator (HE) (Scofield and Kuligowski, 2003), and (6) Precipitation Estimation from Remotely Sensed Imagery using an Artificial Neural Network Cloud Calcification System (PERSIANN-CCS) (Hong et al., 2004).

The HE (Scofield and Kuligowski, 2003), is a numerical weather prediction and brightness temperature-based algorithm. It has undergone a validation over PR by Cruz-González (2006), in which she found that the HE performs better when rain rates are accumulated over longer periods of time than it does for instantaneous readings. She also found that the HE tends to overestimate rain rates over PR. However, another validation performed over PR (Ramírez-Beltrán et al., 2008a), reveals that the HE shows underestimation over a single storm, and a more comprehensive validation results (Ramírez-Beltrán et al., 2008b) showed that the HE is an inconsistent estimator over Puerto Rico. It should be noted that the inconsistency of the HE over Puerto Rico may be attributed to the fact that it was calibrated to operate over the continental U.S. As a result of similar issues encountered with thermal based algorithms different methods of deriving rainfall have emerged. In 2001, Ba and Gruber (2001) presented the GOES Multispectral Rainfall Algorithm (GMSRA), an algorithm that utilizes data from 5 channels of the GOES satellite (i.e., visible (0.65µm), near infrared (3.9µm), water vapor (6.9µm), and window channels (11 and 12µm)).

*Respectively, Graduate Student, Department of Electrical and Computer Engineering, Professor, Department of Industrial Engineering, Professor, Department of Agricultural and Biosystems Engineering, University of Puerto Rico, Mayagüez, PR 00680. E-mails: cardonam@gmail.com, nazario@ece.uprm.edu, eric.harmsen@upr.com.

Recently, two algorithms have been developed that utilize ANN to provide rain rates. These algorithms are the PERSIANN (Sorooshian et al., 2000) and PERSIANN-CCS (Hong et al., 2004). Both algorithms utilize clustering methods (i.e., Kohonen Self Organizing Feature Map (SOFM) (Kohonen, 1982)). In terms of level of extracted information, PERSIANN extracts information at the local pixel scale. PERSIANN-CCS first creates patches of clouds using what the authors called, the Incremental Threshold Temperature algorithm (ITT) (Hong et al., 2004). ITT grows regions of clouds based on a temperature threshold. The authors argue that because of the comprehensive cloud-patch features and the ability to address the variability of rainfall distribution in different cloud clusters PERSIAN-CCS outperforms PERSIANN (Hong et al., 2004).

The major contributions of this study are: (1) creation a cloud classification system, (2) identification of remotely sensed variables that may improve detection and estimation of rainfall, especially rainfall produced by clouds with brightness temperature over 235K, (3) development of a new algorithm to provide rain rates estimates every 15 minutes, (4) development of a reliable product that can be used by the community and/or as input for other models (e.g., flash flood models), and (5) improvement of existing rainfall estimation over PR by the implementation of a cloud classification system and the use of Artificial Neural Networks.

This paper aims to present the efforts made in the development and validation of the cloud classification system and the new rainfall estimation algorithm under development. The next section will offer detailed information on the methodology utilized for this study, followed by preliminary results and conclusions, and finally a summary.

TECHNICAL APPROACH

To meet the objectives established, this study was divided in two major phases. The first phase consisted of the improvement of rainfall detection. During this phase a cloud classification system was developed and the remotely sensed variables that may improve rainfall detection and estimation were identified. After variables had been selected and the classification system developed, the second phase was initiated which consisted of the development of a rainfall estimation algorithm utilizing ANN and the validation of the algorithm in comparison with the HE.

Classification System and Variable Selection

Brightness temperature-based algorithms may fail to detect rainfall produced by clouds with warm tops, because they use a temperature threshold to screen out non-raining clouds. This is the case for the HE which has a brightness temperature threshold of 235K and which fails to capture rainfall events above this temperature in PR (Ramírez-Beltrán et al., 2008a). Researchers have found that other remotely sensed variables such as the effective radius of cloud particles from optically thick clouds can improve rainfall detection during the day-time (Ba and Gruber, 2001).

An array of variables can be derived from GOES-12 data. Some of these variables are: (1) Visual Reflectance centered at 0.65µm, (2) Albedo centered at 3.9µm, (3) Brightness Temperature (Tb) centered at 3.9µm, 6.9 µm, and 10.7, and (4) Brightness Temperature Differences (BTD) between bands (e.g., $Tb_{3.9\mu m}$-$Tb_{10.7\mu m}$, and $Tb_{6.9\mu m}$-$Tb_{10.7\mu m}$). These are shown in Figure 1. In order to identify which remotely sensed variable(s) may help to improve detection and estimation over PR, a classification system was designed. The classification system consists of performing a supervised classification utilizing the maximum likelihood method (MAL). The MAL assumes that the variables follow a Multivariate Normal Distribution; and therefore, the likelihood function to be maximized is given by Eq. 1, where $g_i(x_m)$ is the likelihood of x_m belonging to the class i, x_m is a vector of features to be classified, μ_i is the centroid of the class i, l is the number of features associated with the vector x, S_i is the covariance matrix of the class i and W_i represents the list of variables in class i. Four classes (Figure 2) where previously defined utilizing rainfall estimates from the National Weather Service's (NWS) NEXRAD radar: (1) Rainy clouds with Tb ≤ 235K, (2) Rainy clouds with Tb > 235K, (3) Non rainy clouds with Tb ≤ 235K, and (4) Non rainy clouds with Tb > 235K. These classes were used to train the classifier (i.e., for each class a centroid was computed).

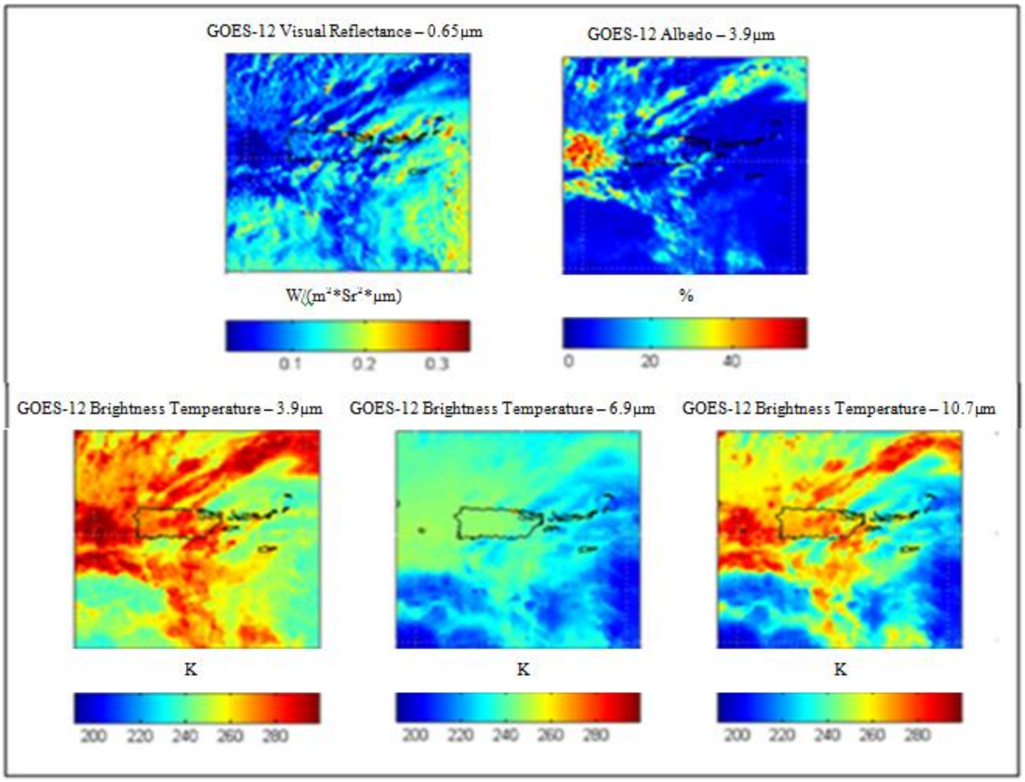

Figure 1. Features extracted from GOES-12 in the vicinity of Puerto Rico.

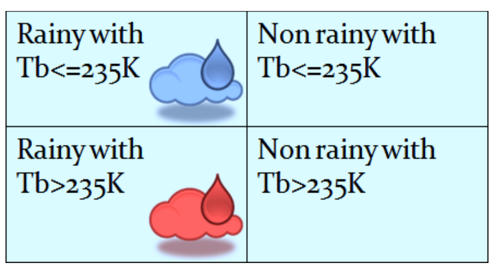

Figure 2. Classes of clouds defined with NEXRAD.

$$g_i(x_m) = \frac{1}{(2\pi)^{\frac{l}{2}} |S_i|^{\frac{1}{2}}} e^{\left[-\frac{1}{2}(x_m-\mu_i)^T S_i^{-1}(x_m-\mu_i)\right]}; \; x_m \in W_i \;\; if \;\; g_i(x_m) > g_j(x_m) \;\; for\ all\ j \neq i \qquad (1)$$

In order to select the variables that may help to improve detection, the classifier's discrete performance (detection) was measured based on hit rate (HIT), bias, probability of detection (POD), and false alarm rate (FAR). For this the classifier was run for 5 storms, utilizing diverse feature vectors, each vector with a different combination of features (i.e., [$Tb_{3.9\mu m}$-$Tb_{10.7\mu m}$, $Tb_{6.9\mu m}$-$Tb_{10.7\mu m}$], [$Tb_{3.9\mu m}$-$Tb_{10.7\mu m}$, Albedo $Tb_{3.9\mu m}$], [$Tb_{3.9\mu m}$-$Tb_{10.7\mu m}$, $Tb_{10.7\mu m}$], etc.). Once the discrete performance was computed for all the runs, a performance index (Eq. 2) defined by Ramirez-Beltran et al., (2009) was utilized to decide which of the feature combinations work best. The index takes into account the FAR, POD and HIT in order to obtain an overall performance of the run. The index will be 0 for a perfect performance (FAR=0, POD=1, HIT=1) and 1 for the worst case scenario (FAR=1, POD=0, HIT=0).

$$index = \frac{FAR - POD - HIT + 2}{3} \quad (2)$$

Rainfall Algorithm Development and Validation

After the variable(s) that may possibly improve detection and estimation of rainfall is identified, a GOES Multi-Spectral Cloud Classification System and ANN Rainfall Algorithm are developed. The algorithm consists of the classification system explained in the previous section for detection improvements, plus two Backpropagation Artificial Neural Networks, one for the rainy clouds with Tb $\leq$ 235K and the other one for the rainy clouds with Tb > 235K, for estimation improvements. Both ANNs are trained with the Levenberg-Marquardt training algorithm. The structure (i.e., initial points, number of neurons in the hidden layer, and transfer function of the hidden layer) of each ANN is selected before the ANN undergoes the training process. In order to select the number of neurons and transfer function in the hidden layer a loop runs with 25% of the data. In the loop two identical neural networks are trained with different amounts of neurons in the hidden layer. One neural network has the Log Sigmoid transfer function in the hidden layer; the other one has the hyperbolic tangent sigmoid. Each neural network is trained with neurons ranging from 2 neurons to 10 neurons in the hidden layer. The number of neurons and transfer function are selected from the neural network with the least error. Once the number of the neurons and the transfer function are selected, a new neural network is created with the selected number of neurons and the selected transfer function.

In order to select the optimal initial point, 50 initial points are tested for the new neural network. As done in the neuron quantity and transfer function selection, in this step the initial point from the neural network with the least error is selected. Finally, once the number of neurons, the transfer function and the initial point are selected, a final neural network is created with the best structure. It is trained with the same portion of data used for the selection of the optimal initial point. After the final neural network has finished it's training, it is used to perform a simulation which is then evaluated to obtain the final performance of the neural network.

Once the algorithm has been fully developed, it will be validated with data from the NWS NEXRAD, located in Cayey, PR. For the validation, diverse heavy rain events registered over PR will be selected. Results will be compared with NOAA/NESDIS operational version of the HE. Both algorithms will be evaluated on a discrete and continuous manner. The discrete validation will measure the capability of the resulting algorithm to detect rainfall. The continuous validation will measure the capability of the resulting algorithm to estimate rain rates.

RESULTS AND CONCLUSIONS

Five heavy storms which occurred during 2003 to 2007 were selected for a discrete comparison between the HE and the cloud classification system in terms of rainfall detection and to select the best feature combinations for rainfall detection. The performance of each method was measured in terms of probability of detection (POD), false alarm rate (FAR), hit rate (HIT), bias and overall performance (INDEX). For this exercise NEXRAD was utilized as ground truth to discriminate between rain/no rain pixels. Table 1 presents the results of the performance of each method for each storm. From Table 1 it can be seen that the cloud classification system outperformed the HE in 4 out of the 5 storms evaluated (lower index value is better). It is important to state that the performance shown for the cloud classification system (CC) belongs to the best feature combination selected (Table 2). Figure 3 shows the rainy pixel detection of the HE algorithm, the CC system and the NEXRAD. This figure shows that the HE overestimates the rainy pixels; whereas, the CC system exhibits an estimation that resembles the NEXRAD detection.

In order to select the best feature combination for rainfall detection with the cloud classification system, an average for the index performance (Eq. 2) was computed for each feature combination for all the storms. The remotely sensed features selected with the best overall performance was the combination of Visual Reflectance centered at $0.65_{\mu m}$, brightness temperature difference between channels 2 ($3.9_{\mu m}$) and 4 ($10.7_{\mu m}$), brightness temperature from channel 4 ($10.7_{\mu m}$) and Albedo centered at $3.9_{\mu m}$. Results presented in this study are associated with daytime periods and in the future the nighttime period will be included.

Table 1. Cloud classification system and HE detection performance comparison.

Storm Date	Method	HIT	POD	FAR	BIAS	INDEX
November - 2003	CC	0.74	0.78	0.15	0.92	**0.21**
	HE	0.53	0.88	0.60	2.30	0.40
December - 2003	CC	0.46	0.30	0.75	1.18	0.66
	HE	0.46	0.50	0.63	3.27	**0.56**
April - 2005	CC	0.65	0.38	0.49	0.74	**0.49**
	HE	0.78	0.15	0.76	0.75	0.61
May - 2005	CC	0.64	0.06	0.46	0.12	**0.59**
	HE	0.46	0.19	0.86	1.34	0.74
October - 2007	CC	0.75	0.80	0.15	0.93	**0.20**
	HE	0.48	0.91	0.57	2.30	0.39

Table 2. Average performance for the best features for rainfall detection.

Feature Combination	$\overline{INDEX}$
$VisRef_{0.65\mu m}$, $Tb_{3.9\mu m}$-$Tb_{10.7\mu m}$, $Tb_{10.7\mu m}$, $Alb_{3.9\mu m}$	0.433
$VisRef_{0.65\mu m}$, $Tb_{6.9\mu m}$-$Tb_{10.7\mu m}$	0.446
$VisRef_{0.65\mu m}$, $Alb_{3.9\mu m}$	0.448
$VisRef_{0.65\mu m}$, $Tb_{3.9\mu m}$-$Tb_{10.7\mu m}$, $Tb_{6.9\mu m}$-$Tb_{10.7\mu m}$, $Tb_{10.7\mu m}$, $Alb_{3.9\mu m}$	0.453
$VisRef_{0.65\mu m}$, $Tb_{6.9\mu m}$-$Tb_{10.7\mu m}$, $Alb_{3.9\mu m}$	0.464

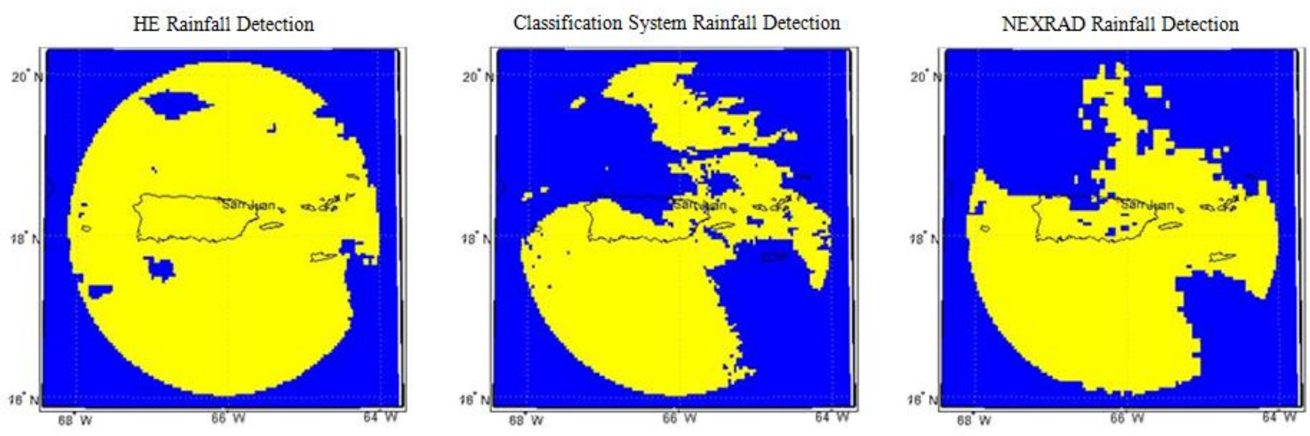

Figure 3. Rainfall detection comparison between the HE (left), the cloud classification system (center), and NEXRAD (right) for October 28, 2007 at 1815 UTC.

SUMMARY

Over the years numerous algorithms to estimate rainfall from remotely sensed data have been developed in order to provide rainfall estimates for those areas where rain gauges are scarce. Some of these algorithms (e.g., NOAA's Hydro-Estimator) are based on a brightness temperature threshold to discriminate from rainy or non-rainy clouds. This method tends to miss rain produced by warm clouds (i.e., clouds with brightness temperature above 235K). This is the case for the HE, which has exhibited a low performance for rainfall detection and estimation in validations performed over PR. For these reasons researchers have develop new and diverse methods to remotely detect and estimate rainfall (e.g., GMSRA, PERSIANN, and PERSIANN-CCS).

In this paper a new method to detect and estimate rainfall combining diverse features obtained from GOES-12, a cloud classification system and artificial neural networks was presented. In terms of detection, the new algorithm outperformed the HE in 4 out of 5 storms. This offers evidence that the algorithm represents a potential improvement over the HE for PR. However, more work is needed in order to validate the algorithm and compare the performance with the HE in terms of rainfall estimation.

ACKNOWLEDGMENTS

This work is supported primarily by NOAA-CREST (grant NA06OAR4810162) and the Engineering Research Centers Program of the National Science Foundation (grant 0313747). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the NOAA and NSF.

REFERENCES

Arkin, P. A., and Meisner, B. N., 1987. The Relationship Between Large Scale Convective Rainfall and Cold Cloud Over the Western Hemisphere During 1982–84, Monthly Weather Review, Vol. 115, pp. 51-74.

Ba, M. B., and Gruber, A., 2000. GOES Multispectral Rainfall Algorithm (GMSRA), Journal of Applied Meteorology, Vol. 40, pp. 1500-1514.

Cruz-González, B., 2006. Validación del Algoritmo Hidro-Estimador en la Region de Puerto Rico, MS thesis, Industrial Engineering Department, University of Puerto Rico, Mayagüez, Campus.

Hong, Y., Hsu, K. L., Sorooshian, S., and Gao, X. G., 2004. Precipitation Estimation from Remotely Sensed Imagery Using an Artificial Neural Network Cloud Classification System, Journal of Applied Meteorology, Vol. 43, No. 12, pp. 1834-1852.

Kohonen, T., 1982. Self-Organized Formation of Topologically Correct Feature Maps, Biological Cybernetics, Vol. 43, pp. 59-69.

Ramírez-Beltrán, N. D, Kuligowski, R. J., Harmsen, E. W., Castro, J. M., Cruz-Pol, S., Cardona-Soto, M. J., 2008a. Rainfall Estimation from Convective Storms Using the Hydro-Estimator and NEXRAD, WSEAS Transaction on Systems, Vol. 7, No. 10, pp. 1016-1027.

Ramírez-Beltrán, N. D. and Cruz-Pol, S., 2008b. Statistical Techniques to Improve the Hydro-Estimator Rainfall Algorithm During Heavy Storms over Puerto Rico, Final report, NOAA, November, 2008. Grant Number: NA06NWS4680011.

Ramírez-Beltran, N.D, Kuligowski, R.J., Castro, J.M.,Cardona-Soto, M, Vasquez, R., 2009. A Projection Algorithm for Satellite Rainfall Detection. *WSEAS Transaction on Systems*. Issue 6, Vol 8, pp 763-772. http://www.worldses.org/journals/systems/systems-2009.htm

Scofield R. A. and Kuligowski R. J., 2003. Status and Outlook of Operational Satellite Precipitation Algorithms for Extreme-Precipitation Events, Weather and Forecasting, Vol. 18, No. 6, pp. 1037-105.

Sorooshian, S., Hsu, K.-L., Gao, X., Gupta, H. V., Imam, B., and Braithwaite, D., 2000. An Evaluation of PERSIANN System Satellite-Based Estimates of Tropical Rainfall, Bulletin of the American Meteorological Society, Vol. 81, pp. 2035-2046.

Vicente, G., Scofield, R. A., and Mensel, W. P., 1998. The Operational GOES Infrared Rainfall Estimation Technique, Bulletin of the American Meteorological Society, Vol. 79, pp. 1881-1898.

Vila D. and Velasco I., 2002. Some Experiences On Satellite Rainfall Estimation Over South America, 1st IPWG Workshop EUMETSAT SAF NWC, Instituto Nacional de Meteorología, Madrid, Spain, 23-27 September.

Vila D. A., Scofield, R. A. Kuligowski, R. J., Davenport, J. C., 2003. Satellite Rainfall Estimation Over South America: Evaluation Of Two Major Events, NOAA Technical Report NESDIS Nº 114, U.S. Department of Commerce, pp. 1-17.